Decision Graphs from ReLU Sign Patterns using Walsh–Hadamard Commutator Operators

Introduction

Conventional ReLU neural networks can be viewed as dynamically selecting submatrices of a weight matrix through binary activation decisions. Normally this selection process is represented implicitly:

$$y = W\,\mathrm{ReLU}(x)$$

or equivalently:

$$y = WD(x)x$$

where $D(x)$ is a diagonal matrix containing 0 or 1 entries depending on activation sign.

This interpretation is useful because it reveals that a ReLU network is not using one fixed linear operator. Instead, each input generates a different effective matrix through the activation mask $D(x)$.

However, ordinary ReLU masking has several limitations:

- information is destroyed when activations are zeroed,
- rank is often reduced,
- routing decisions are coordinate-local,
- and interactions between channels are not explicitly represented.

An alternative approach is to preserve sign information using diagonal matrices with entries ±1 instead of 0/1. In this formulation the activation pattern does not suppress channels outright. Instead it encodes a binary partition of feature space.

When combined with a structured orthogonal mixing transform such as a Walsh–Hadamard matrix H, this produces much richer dynamic operators:

HD + DH

And

HD - DH

These matrices behave like dynamically generated graph operators whose structure depends on the activation signs.

ReLU Networks as Decision Matrix Systems

A standard ReLU layer may be rewritten as:

y = WDx

where:

- D is diagonal,
- D_{ii} = 1 for active units,
- D_{ii} = 0 for inactive units.

This makes it clear that ReLU performs dynamic feature selection.

The network effectively chooses:

- which rows participate,
- which columns contribute,
- and which paths through the network are active.

From this viewpoint, deep ReLU networks are systems which compose large numbers of diagonal decision matrices with learned dense matrices.

The main limitation is that these decisions are independent per coordinate. The mask only determines whether a feature exists or not.

No explicit relational structure between features is represented.

Replacing 0/1 Decisions with ±1 Decisions

Suppose instead that the diagonal matrix contains:

- +1 for positive activations,
- −1 for negative activations.

Then:

$D_{ii} \in \{-1,+1\}$

The activation pattern now partitions channels into two groups rather than suppressing half the state space.

This has several consequences:

- sign information is preserved,
- norm loss is reduced,
- routing becomes phase-like rather than destructive,
- and orthogonal transforms remain energy-preserving.

Most importantly, the sign pattern can now be used to generate interaction topology.

Walsh–Hadamard Mixing

Let H be a Walsh–Hadamard matrix.

Walsh–Hadamard transforms are attractive because they are:

- orthogonal,
- extremely fast,
- binary structured,
- fully mixing,
- and parity-sensitive.

Applying H globally mixes all channels using only ±1 operations.

Conjugating diagonal sign matrices through H produces dense structured operators:

HDH

which can be interpreted as transformed interaction graphs.

The combinations:

HD + DH

and

HD - DH

become particularly interesting.

Emergence of +2, 0, and −2 Structure

Consider entries of:

HD + DH

For matrix element ij:

$(HD + DH)_{ij} = H_{ij}(D_{jj} + D_{ii})$

Since D_{ii} and D_{jj} are each ±1, their sum can only be:

- +2,
- 0,
- or −2.

Similarly:

$(HD - DH)_{ij} = H_{ij}(D_{jj} - D_{ii})$

which also produces:

- +2,
- 0,
- or −2.

This creates structured ternary interaction matrices.

The meaning is important.

For HD + DH:

- equal-sign node pairs survive,
- opposite-sign pairs cancel.

For HD − DH:

- opposite-sign node pairs survive,
- equal-sign pairs cancel.

Thus the activation pattern generates two complementary interaction graphs:

Operator	**Interpretation**
HD + DH	coherence graph
HD − DH	boundary/disagreement graph

The zeros are not random sparsity. They arise from relational cancellation conditions between pairs of features.

Graph Interpretation

The sign vector encoded in D partitions channels into positive and negative groups.

The operator HD + DH connects channels with matching sign state.

The operator HD − DH connects channels with opposing sign state.

This resembles graph constructions used in:

- signed graphs,
- graph cuts,
- Ising systems,
- spectral clustering,
- and parity networks.

The resulting matrices are dynamic graph operators generated directly from activation structure.

Unlike ordinary ReLU masks, these operators do not merely suppress coordinates. They alter pairwise interaction topology.

Neural Layer Construction

A layer can then be written as:

$$y = (HD + DH)W_1x + (HD - DH)W_2x$$

where:

- W_1 processes coherent interactions,
- W_2 processes disagreement interactions.

The layer decomposes computation into:

- within-partition communication,
- and cross-partition communication.

This is substantially richer than ordinary coordinate gating.

The activation pattern dynamically defines which interactions exist.

Relation to Attention Mechanisms

This construction resembles attention in that pairwise interactions depend on the input.

However, unlike attention:

- no dense similarity matrix is computed,
- no quadratic softmax is required,
- and the structure is algebraically constrained.

The routing is generated through:

- diagonal sign flips,
- Walsh mixing,
- and addition/subtraction.

This gives a form of structured dynamic connectivity at very low computational cost.

Information Preservation

Ordinary ReLU networks discard negative information entirely.

In contrast, ±1 decision systems preserve sign state and use it constructively.

This changes the role of activations:

Conventional ReLU	**±1 Sign System**
suppress features	encode phase state
reduce dimensionality	preserve dimensionality
local gating	relational gating
coordinate masking	graph construction

The system becomes less like sparse thresholding and more like dynamic interaction geometry.

Symmetric and Antisymmetric Dynamics

The two operators have qualitatively different behavior.

HD + DH is symmetric-like and tends to reinforce coherent structure.

HD − DH is antisymmetric-like and emphasizes transitions and boundaries.

This resembles decompositions appearing in:

- diffusion versus circulation,
- conservative versus rotational flow,
- agreement versus disagreement dynamics.

The antisymmetric branch may support oscillatory or rotational internal representations, while the symmetric branch may stabilize coherent feature clusters.

Iterated Systems

Repeated application may generate highly structured dynamics:

$$x(t+1) = (HD \pm DH)x(t)$$

Because H is orthogonal and D only changes signs, the system may preserve substantially more information than conventional gated layers.

Deep compositions such as:

$$HD_1HD_2HD_3$$

or mixed forms such as:

$$(HD_1 - D_1H)(HD_2 + D_2H)$$

effectively compose evolving interaction graphs.

This begins to resemble programmable graph algebra generated directly from activation patterns.

Potential Advantages

Several possible advantages emerge:

- lower information loss,
- structured dynamic sparsity,
- efficient conditional routing,
- richer feature interactions,
- parity-sensitive computation,
- and graph-like relational processing.

Because Walsh transforms are computationally cheap, these mechanisms may scale efficiently compared to learned dense attention systems.

Open Questions

Several areas deserve investigation:

- spectral stability of repeated compositions,
- gradient propagation through sign structures,
- normalization requirements,
- representational rank preservation,
- emergent graph geometry,
- and learning behavior of symmetric versus antisymmetric branches.

It is also unclear whether learned Walsh-like orthogonal transforms may outperform fixed Hadamard structure in this framework.

Conclusion

Viewing ReLU activations as decision matrices suggests a broader interpretation of neural computation as dynamic operator selection.

Replacing binary 0/1 masks with ±1 sign structures allows activation patterns to generate relational interaction graphs rather than merely suppressing coordinates.

The operators:

HD + DH

and

HD - DH

naturally produce ternary interaction structure with entries in {+2, 0, −2}. These correspond to coherent and boundary interactions between activation groups.

This transforms activation gating into a form of dynamic graph construction driven directly by sign topology.

The result is a computationally cheap but structurally rich system combining:

- orthogonal mixing,
- parity structure,
- graph partition dynamics,
- and conditional interaction routing.

Such systems may provide an alternative path toward low-information-loss neural architectures with richer internal relational structure than conventional ReLU networks.